GENERALIZED EPIPHENOMENALISM

I want to show that a common and plausible interpretation of what science tells us about the fundamental structure of the world – the ‘scientific picture of the world’ or SPW for short – leads to what I’ll call ‘generalized epiphenomenalism’, which is the view that the only features of the world that possess causal efficacy are fundamental physical features. I think that generalized epiphenomenalism follows pretty straightforwardly from the SPW as I’ll present it, but it might seem that, once granted, generalized epiphenomenalism is fairly innocuous, since its threat is too diffuse to provoke traditional worries about the epiphenomenal nature of mental states. If mental states are epiphenomenal only in the same sense that the putative powers of hurricanes, psychedelic drugs or hydrogen bombs are epiphenomenal, then probably there is not much to worry about in the epiphenomenalism of the mental. I agree that the epiphenomenalism of hurricanes and the like is manageable, but it will turn out that ensuring manageability requires that mental states have an ontological status fundamentally different from that of hurricanes, drugs and bombs, a status that is in fact inconsistent with the SPW. So I’ll argue that generalized epiphenomenalism does have some seriously worrying consequences after all.

The SPW takes as its starting point the modern naturalistic conviction that the basic structure of the world can be discovered by scientific investigation and that there is no ground for positing a metaphysical understanding of the world distinct from a scientific understanding (a slogan: fundamental science is just metaphysics with numbers). Three interlocking features seem of central importance to the SPW: completeness, closure and resolution. Completeness is the doctrine that everything in the world is physical and as such abides by closure and resolution. Closure entails that there are no ‘outside forces’ – everything that happens, happens in accordance with fundamental physical laws so as to comply with resolution. Resolution requires that every process or object be resolvable into elementary constituents which are, by completeness, physical and whose abidance with laws governing these constituents leads to closure[1].

[1] It may be worth noting here that this is not an endorsement of so-called ‘part-whole reductionism’, though it is consistent with it. We know from quantum mechanics that the states of ‘wholes’ are not simple functions of the states of their parts but this does not tell against the characterization given in the text. Quantum mechanics is a *celebration* of how the interactions of things can be understood – rigorously understood – to yield new features. It is, if you like, the mathematical theory of emergence, but one that obeys the strictures of resolution.

Take anything you like: a galaxy, a person, a flounder, an atom, an economy – it seems that anything can be *resolved* into the fundamental physical constituents, processes and events which determine its activity. Indeed, our best theory of the creation of the universe maintains that at very early times after the creation event (the 'big bang') the universe was quite literally resolved into its elementary constituents; at that time the universe consisted of an extremely hot, highly active 'gas' of quarks (that would later, upon sufficient cooling, combine into more familiar composite particles such as protons and neutrons), leptons (including the electron needed for the later – after still more cooling – transformation of protons and neutrons into atoms, forming the basis of *chemical* kinds) and elementary force exchange bosons (whose interactions with the other particles, roughly speaking, provide the 'go' – to use the nice term of the emergentist C. Lloyd Morgan – of the universe, driving it from state to state)[2]. Completeness, closure and resolution and their inter-relations are concisely expressed in the thought that the universe is 'running' entirely and solely upon the interactions of these elementary constituents no less today than when it was 10^{-37} seconds old.

It is crucial to emphasize that the SPW is a metaphysical, not an epistemological doctrine. It does not concern itself with how or whether we could understand everything in terms of full resolution. In fact, such understanding is quite impossible, for reasons of complexity (of various sorts) that the SPW itself can spell out. Innumerable immensely difficult questions arise at every stage of resolution[3], and there is no practical prospect whatsoever of *knowing* the full details of the physical resolution of anything much more complex than even a simple atom. The metaphysical picture is nonetheless clear. And since the *world* has no need to know the details but just runs along *because* the details are the way they are, the problems *we* have understanding complex systems in terms of fundamental physics are quite irrelevant to the metaphysics of the SPW. The

[2] This is expressed from a 'particle perspective'. All these particles are quanta of underlying quantum fields that perhaps have a better claim to be the fundamental physical structure of the world. Of course, there is no guarantee that the quarks, leptons and bosons are truly elementary; perhaps they are themselves composites of ultra-microscopic 'strings'.

[3] For example, something as simple as the spin of a proton turns out to be the product of an almost inconceivably complex interaction between the three constituting (or 'valence') quarks, a ghostly host of virtual particles within the proton as well as an additional component of orbital spin. See ??.

SPW maintains that the world resolves itself into fundamental entities whose properties determine the entire structure of the world and that it is the job of physics to provide an absolutely exact representation of these entities and their basic interactions.

There is a kind of philosophical model which reveals how completeness, closure and resolution are supposed to work, and which might also serve as a test of one's attitude towards the SPW. Call the model the superduper computer simulation thought experiment. It goes like this.

Imagine the day when physics is complete. A theory is in place which provides an exact representation of the truly elementary entities and processes which form the physical, and metaphysical, foundation of the world. There are some who see this day as perhaps not too distant (e.g. Hawking 1988, Weinberg 1992, Horgan 1996). Of course, the mere possession of this *theory* of everything will not give us the ability to provide a complete *explanation* of everything: every event, process, occurrence and structure. Most things will be too remote from the basic theory to admit of explanation in its terms; even relatively small and simple systems will be far too complex to be intelligibly described in the final theory.

But seeing as our imagined theory is fully developed and mathematically complete it will enable us to set up detailed computer simulations of physical systems. The range of practicable simulations will in fact be subject to the same constraints facing the explanatory use of the theory; the modelling of even very simple systems will require impossibly large amounts of computational resources. Nonetheless, possession of a computational implementation of our final theory would be immensely useful. Real versions of something very like my imaginary scenario now exist and are already fruitful. For example, there are computer models of quantum chromodynamics that can compute the theoretically predicted masses of various quark constituted sub-atomic particles (see Weingarten 1996). The looming problem of computational intractability is all too evident, for realizing these calculations required the development of special mathematical techniques, the assembling of a dedicated, parallel supercomputer specially designed for the necessary sorts of calculations (a computer capable of 11 billion arithmetical operations per second) and roughly a *year* of continuous computing. Weingarten reports that a special 2-year calculation revealed the existence of a previously unrecognized particle, whose existence could be verified by examining past records from particle accelerator experiments. Modelling the *interactions* of particles would be a much more challenging task, suggesting to the imagination computational projects analogous to the construction of medieval cathedrals, involving thousands of workers for many decades.[4]

Now I want to introduce a thought experiment that flatly ignores the inevitably insuperable problems of computational reality. Imagine a computer model of the final physical theory which has no computational limits (we can deploy as much memory as we like and compute for as long as we like): imagine that detailed specifications of the basic physical configuration of any system, at any time, in terms appropriate for the final theory, are available, so that if the configuration of any physical system is specified as input then the output configuration of the system, for any later time, can be calculated (and appropriately displayed). If the final theory should turn out to be non-deterministic (unlikely as that may seem, given that quantum mechanics, which seems to form the basis of any final physics now envisageable, provides for the completely deterministic evolution of the wave function of any system[5]) then we can permit multiple simulations to run simultaneously, thus to duplicate the statistics to be found in the real world. There is nothing incoherent in the idea of an absolutely perfect simulation. In fact, we might have some of them in physics already. The Kerr equations for rotating black holes are (if the general theory of relativity is true), absolutely perfect models of these strange objects. The famous astrophysicist and mathematician Subrahmanyan Chandrasekhar confessed that ‘in my entire scientific life ... the most shattering experience has been the realization that an exact solution of Einstein’s equations of general relativity, discovered by the New Zealand mathematician Roy Kerr, provides the

---

[4]Would it ever make sense to start such a project? Not if computer technology progresses sufficiently quickly. Suppose the original length of the computation is *n* years and technology advances so quickly that after *d* years have passed the computation would take less than $n - d$ years. Then it would never make sense to start the computation! Of course there are non-computer technical constraints on the time required for such computations and presumably the pace of progress in computer technology must eventually slow down rather than continue its heretofore exponential acceleration. For some *n*, the computations make sense, as evidenced by the real world examples given above, but the problem of this note is equally well illustrated (see Weingarten 1996). Computers of the 1980s would have taken about 100 years to perform the reported computations. It was not worth starting.

[5] There are very large issues lurking here, most especially the so-called measurement problem. If one believes that there are non-deterministic processes that drastically and uncontrollably alter the wave function then we can resort to the multiple simultaneous simulation model outlined immediately below.

absolutely exact representation of untold numbers of massive black holes that populate the universe' (as quoted in Begelman and Rees 199?, p. 188).

However, we are not so lucky with the rest of the world, and so even within our dream certain approximations in the input configurations will have to be allowed. We cannot input field values for every point of space-time and it is conceivable that some configurations require an infinite amount of information for their specification if, to give one example, certain parameters take on irrational values which never cancel out during calculation. Let us therefore imagine that we can input specifications of whatever precision we like, to allow for modelling the system for whatever time we like, to whatever level of accuracy we desire. Even though it is not physically realizable, I think the idea of such a computer program is perfectly well defined.[6]

So, let us imagine a computer simulation of a part of the world. Step one is to restrict our attention to something we call 'simple' – a bob on a spring on the moon say. The simulation covers a restricted region of space and time (though the programmer can set up 'boundary conditions' that represent the influence of the rest of the world), and must be defined solely in terms of the values of *fundamental* physical attributes over that region. The programmer is not allowed to work with gross parameters such as the mass of the bob or the stiffness of the spring, or the gravitational force of the moon, but must write her code in terms of the really basic physical entities involved. (It might help to imagine the code written in terms of the properties of the atoms of the pendulum, its support structure and moon, though these are themselves not really physically basic.) The SPW predicts that the output of this computer simulation, appropriately displayed, would reveal a bob bouncing up and down, suspended above the lunar surface. Step two is to up the ante. Now imagine a simulation of a more complex situation, for example a father and child washing their dog, in their backyard on a lovely sunny day. Do you think the simulation would mimic the actual events? I venture to maintain that what we know about the world very strongly suggests that such a simulation would 're-generate' both the action of the pendulum and the

---

[6]Could we define physicalism in terms of this imaginary computer implementation of final physics? We might try something like this: physicalism is the doctrine that everything that occurs/exists in the actual world would have its exact counterpart in a final physics computer simulation of the world, or that the simulation would be, in some appropriate sense, indistinguishable from the actual world. Such a formulation has the advantage of automatically including what Hellman and Thompson (1975) call the principle of physical exhaustion. But it obviously requires a clearer specification.

behaviour of the father, child and dog (along with tub, water, soap, sunlight, etc.).

Although something of a digression, it is worth considering the details of such simulations a little more closely. The thought experiment is outrageously idealised. We assume unlimited (but finite) memory and allow unlimited (but finite) processing time. Even relative to such generous constraints, there are questions about the ‘feasibility’ of the simulation. The field of applied mathematics that deals with numerical approximations of differential equations is exceedingly complex and it is not clear that all physical systems (describable by differential equations) can be accurately simulated. It is known that not all such systems can be simulated if we require the simulation to obey certain otherwise apparently desirable mathematical constraints (see Umeno, Ge and Marsden ??). Although I assume that the mathematics of the eventual final physics will be simulatable in the weak sense required here, how likely this really is I have no idea.

It is also possible that nature transcends the ‘Turing Limit’ – that is, can only be described in terms of uncomputable functions. Thus, it is not known with certainty that a digital computer (equivalent to a universal Turing machine) can simulate all the mathematical functions which describe nature. Nature might use, so to speak, uncomputable functions in getting the world to move from state to state. A very simple example (from Copeland ??) of an uncomputable function is the function $E(x, y)$ defined as $E(x, y) = 1$ if $x = y$ and $E(x, y) = 0$ if $x \neq y$. Even if x and y range over computable real numbers, E is not computable (since you’d have to check an infinite number of digits to verify the identity of two real numbers). If nature ‘uses’ real numbers in ways something like the function E then our simulation may turn out to be impossible.

But notice that we don’t necessarily have to restrict ourselves to Turing machines (or the souped up equivalents we all have on our desks). If nature works ‘beyond the Turing limit’ then perhaps we can build computers that exploit that region of physics. One example of such an ‘enhanced’ computer that has been studied by theoretically minded computer scientists is what Turing himself called *Oracle* machines. These are Turing machines that can, at well defined points during computation, call upon an ‘oracle’ to provide the output of an uncomputable function. No one knows if oracle machines are constructable, but if nature herself ‘uses’ uncomputable functions then nothing seems to prevent us from in principle incorporating such aspects of nature into our computing machinery. For the argument advanced in this paper, the crucial constraint involves resolution. The simulation must operate solely over the basic constituents of nature and only over the properties of those constituents. The importance of this restriction has to do with the notion of emergence, to be discussed immediately below.

The simulation thought experiment can be used to provide a simple and clear definition of *emergence*. An emergent is anything that is not coded into the simulation. Thus a thunderstorm is an emergent *entity* since, I take it, we would not need, in addition to coding in the quarks, leptons and bosons and their properties, to add thunderstorms as such to our simulation code. Temperature would be an example of an emergent *property* (thermodynamical properties would in general be emergent properties), as would be such features as ‘being hydrogen’ (chemical properties in general would be emergent properties), ‘being alive’ (biological properties would in general be emergent properties), etc. The founders of the doctrine of emergentism (see Lewes, Morgan, Alexander, Broad) would have agreed with my examples as examples of emergent features, but they wanted more from their emergents; they wanted their emergents to have an active role in the go of the world, not a merely passive or derivative role (see MacLauglin for a review of so-called British emergentism’s nature and goals). The world, according to these emergentists, goes differently because of the presence of emergents; it does not behave simply in the way it would if the properties of the basic physical constituents of the world were solely efficacious and the emergents were simply ‘conglomerations’ of basic physical entities. In deference to their desires, we can distinguish a *benign* from a *radical* emergence. In terms of our simulation thought experiment, radical emergence is the claim that the simulation will, despite being an entirely accurate representation of the basic physical constituents of the world, not render an accurate simulation of the development of the world as these basic constituents combine and interact to form ever more complex structures. That is, despite the accuracy of the simulation with regard to basic physical features, we will have to code into our simulation additional features that come into play only when certain *combinations* of the basic features appear[7].

The emergentists’ favourite example of what they took to be a relatively straightforward,

[7] There are some obvious epistemological problems here. How would one know a case of radical emergence from a merely false theory of the basic constituents? One can imagine various ways such difficulties could be addressed. For example, suppose – what we already know to be false – that our best theory of the elementary features could not explicate even the simplest chemical properties of atoms. After enough failure, we might have reason to come to believe that chemical properties were brute and radically emergent features of certain complex structures.

uncontroversial example of radical emergence was chemistry. They took it to be the case that the theory of the atom simply could not account for all the various chemical properties and chemical interactions found in the world. And they meant this to be a metaphysical claim. They were well aware of the difficulties of complexity which stood, and stand, in the way of fully *understanding* chemistry in terms of physics, or feasiblely *predicting* chemical properties on the basis of basic physical properties. The emergentists were not merely noting these difficulties of complexity. They were denying that chemistry *resolves* into physics (or that chemical entities resolve into physical entities) so as to obey *closure*. As Broad put it, even a 'mathematical archangel' could not deduce chemistry from physics. In place of the angel, I've put our simulation of basic physics, so the claim of radical emergentism is just that the simulation will not provide an accurate representation of distinctively chemical events. This is a kind of empirical claim. And though in a certain sense it is untestable, the development of quantum mechanics has severely undercut the case for chemistry being radically emergent. It now seems clear that chemical properties emerge from the properties of the basic physical constituents of atoms and molecules entirely in accord with closure and resolution. And that is the sign of *benign* emergence. Benign emergence is simply the claim that all features not coded into the simulation are subject to resolution under closure.

Thus the SPW, with its endorsement of completeness, closure and resolution, asserts that all emergence is benign emergence. Now, does benign emergence entail generalized epiphenomenalism? The threat is clear. Benignly emergent features have no *distinctive* causal job to do; whatever they 'do' is done through the agency of the basic physical features that subvene them. But that does not directly entail epiphenomenalism. The SPW does not impugn the existence of benignly emergent features so one might suspect that some notion of 'supervenient causation' will suffice to grant efficacy to these emergents. And let me emphasize again, they have indispensable explanatory jobs and there is no prospect of or desire for their elimination from the world.

Nonetheless, I believe that generalized epiphenomenalism does follow from the SPW. There are at least three arguments for this conclusion, which I'll label the *economy argument*, the *screening-off argument* and the *abstraction argument*.

The Economy Argument

I take it that, as mentioned above, causation is a metaphysical relation, and that in particular it is the relation that draws forth one event from another across time, or that determines one state as the outcome of a previous state. This is no definition and I don't want to prejudge issues about the possibility of backwards causation or about whether the relata of the 'causes' relation are necessarily and only events as, for example, opposed to states or objects. I want only to draw attention to the central idea that causation is the go of the world; it is causation that drives the world from state to state. The issue of concern to me is, so to speak, how much go there is in the world or how widely it is 'spread out' through the world.

The definite question I want to address is this: is there, from the point of view of the scientific metaphysics outlined above, any need to posit causal efficacy at any level above that of the fundamental physics, or is all of the go lodged at the metaphysical root of the world? This question must be sharply distinguished from the question whether we need to deploy theories of (or descriptions of) levels of reality far higher than those described by fundamental physics in order to predict occurrences in the world, to explain what happens in the world and to understand or comprehend what is happening around us. I think it is obvious that we require high-level theories or descriptions for these essentially epistemic tasks. You are not going to *understand* why a square peg won't fit in a round hole in terms of the fundamental physics governing the constituents and environs of peg and hole. But that by itself doesn't entail that we need to posit any causal efficacy to 'square pegged-ness' or 'round holed-ness'. No less evident than the need for high-level descriptions to understand this relationship, is that the fundamental physics is all you need to ensure that, as a matter of fact, the square peg just won't go into the round hole.

The superduper computer simulation thought experiment is supposed to draw this to our attention. Imagine the fundamental physics simulation of peg approaching hole. There is no need to code into the simulation anything about *squareness* or *roundness*, or whether something is a *peg* and something else is a *hole*, or that the *peg is moving towards the hole* or anything else at a level of description above that of fundamental physics. Nonetheless the world of the simulation reveals that the peg won't go through the hole. How can that be if there really is some kind of genuine causal efficacy to the peg's being square or the hole's being round? It would seem reasonable to suppose that if you leave some genuine efficacy out of your simulation, it won't manage to remain similar to the world in which that missing efficacy resides and has its effects (recall our definition of ‘radical emergence’ versus ‘benign emergence’ here). Leaving out of the simulation features that make a genuine causal contribution to the evolution of the world’s state ought to cause the simulation to drift out of synchronization with the real world. But, by our hypotheses of completeness, closure and resolution, *no* high-level features are ever needed to get our simulation to accurately duplicate the world.

Of course, if you regard causation as a non-metaphysical relation – perhaps some kind of explicatory notion then I grant you its existence in the high-level features (and, I suppose, it is actually *missing* from the low-level fundamental features that are too complex and particular to explain things like why square pegs won’t go in round holes). It is not unthinkable that our common sense notion of causation is rather unclear about the distinction between epistemology and metaphysics (and this confusion might account for much of the trouble we have making sense of such things as the ‘causal relevance’ of, for example, mental properties). But whatever the proper analysis of ‘causes’ may be, there remains the metaphysical question of where the go of the world resides and how much of it has to be posited to get the world going the way it actually does go.

If you think that causation is primarily an epistemological or explicatory relation, or that both metaphysical and epistemological notions jointly constitute our concept of causation, I won’t argue about the *word*. Define ‘kausation’ as the *metaphysical* relation between events (or whatever) that drives the world forward. Our question then is whether high-level features have any kausal efficacy; the metaphysical question remains as pressing as ever. (In what follows, however, I keep to the ‘c’ spelling.)

We ought not to multiply entities beyond necessity. In the particular case of the metaphysical question of where causation works in the world and how much of it there is, we ought to posit the minimum amount, and the simplest nature, necessary to get the phenomena we seek to account for. The ‘phenomena’ are just the events that make up our world (at any level of description). The minimum amount of causation we need to posit is causation entirely restricted to the level of fundamental physics. This follows from closure, completeness and, especially, resolution. Fundamental physics (at the moment) suggests there are, currently active, four forces (weak, strong, electromagnetic, and gravitational) whose concerted exertions (within a backdrop of spacetime and quantum fields) generate all the variety we can observe in the world at large.

Crudely speaking, our superduper simulation requires only the simulation of these forces (and fields) in spacetime to provide a total simulation of the world at every level.

Consider, as an example of a high-level feature, the coriolis force, which gunnery officers must take into account when computing the trajectory of long-range cannon shells. The general significance of the coriolis force can be gathered from this passage from the *Encyclopedia Britannica*:

> The Coriolis effect has great significance in astrophysics and stellar dynamics, in which it is a controlling factor in the directions of rotation of sunspots. It is also significant in the earth sciences, especially meteorology, physical geology, and oceanography, in that the Earth is a rotating frame of reference, and motions over the surface of the Earth are subject to acceleration from the force indicated. Thus, the Coriolis force figures prominently in studies of the dynamics of the atmosphere, in which it affects prevailing winds and the rotation of storms, and in the hydrosphere, in which it affects the rotation of the oceanic currents. (*Britannica '98*, entry for Coriolis Force)

This is a benignly emergent property of the earth, or any other rotating system, of evident usefulness in a variety of high-level descriptions of the world. But in the context of assessing genuine causal efficacy in terms of the fundamental physical features of the world, it is highly misleading to say that the coriolis force *causes* diversions in, for example, a shell's trajectory. At least, if we really thought there was such a force – hence with its own causal efficacy, the world would end up being a much stranger place than we had imagined. Just think of it: rotate a system and a brand new force magically appears out of nowhere, stop the rotation and the force instantly disappears. That is radical, brute emergence with a vengeance. Of course, there is no need to posit such a force (and it is called, by physicists if not engineers, a fictitious force, as is *centrifugal* force). The coriolis phenomena are related to the underlying physical processes in a reasonably simple way – in fact simple enough for us to 'directly' comprehend, but, no matter the complexity, our imaginary computer model of any rotating system would naturally reveal the appearance of a coriolis force.

The coriolis force can serve as a more general model for the relation between basic and high-level features of the world. The coriolis force is an 'artifact' of the choice of a certain coordinate system. If you insist upon fixing your coordinates to the surface of the Earth, you will notice the coriolis force. If you take a more natural, less geocentric, non-rotating, coordinate system as basic, the force will never appear (though artillery shells will of course still track a curved path across the surface of the Earth). In general, high-level features are ‘artifacts’ which arise from the selection of a particular mode of description. If, so to speak, we impose the ‘chemical coordinate system’ upon ourselves, we will find peculiar ‘chemical forces’ at work (or we will find chemical properties to be apparently efficacious). If we stick with the fundamental basic physical framework, these distinctively chemical activities will as it were disappear, though, of course, the phenomena chemists like to explain via ‘co-valent bonds’, ‘hydrogen bonds’, etc. will still occur.

High-level features thus are in essence what Daniel Dennett has called ‘patterns’ (see Dennett ??). Patterns are structures, and relations amongst structures, that are visible from certain viewpoints or what I have metaphorically labelled co-ordinate systems. Patterns are somewhat odd in that they inhabit a curious zone midway between, as it were, objectivity and subjectivity for patterns are *there* to be seen, but have *no function* if they are not seen[8]. By the former, I mean that patterns are not just in the eye of the beholder; they are really in the world (it’s not *optional* for us to decide that salt does or does not dissolve in water) and they provide us with an indispensably powerful explanatory and predictive grip upon the world. By the latter, I mean that the *only* role they have in the world is to help organize the experience of those conscious beings who invent them and then think in terms of them. That is, although the world is rightly described as exemplifying a host of patterns, the world has no use for them. In terms of our thought experiment again, high-level patterns do not need to be coded into the world-simulation in order to ensure the accuracy of the simulation, and this is just because it is the fundamental physical features of the world which organize the world into all the patterns it exemplifies and they do this all by themselves, with no help from ‘top-down’ causation[9].

---

[8] Note that Dennett says ‘These patterns are objective – they are there to be detected – but from our point of view they are not out there independent of us, since they are patterns composed partly of our ‘subjective’ reactions to what is out there, they are the patterns made to order for our narcissistic concerns’ (1987, p. 39).

[9] I note that Jaegwon Kim attempts to ground the causal efficacy of higher-order features with the claim that ‘the causal powers of an instance of a second-order property are identical with (or a subset of) the causal powers of the first-order realizer that is instantiated on that occasion’ (Kim

Doubtless there is a harmless sense of ‘top-down causation’ which is perfectly acceptable, appropriate for use within pattern-bound explanations. For example, we can explain the location of a particular atom by reference to the *intentions* of the operator of a scanning tunnelling electron microscope. But we know that those very intentions are elusively accommodated within a vastly intricate web of micro-states which, within their environment, ‘push’ the target atom to its final location. Intentions, like planets, animals and molecules, have no need to be specially written into the code of the world-simulation.

Thus, the simplest and most natural interpretation of efficacy in the SPW restricts efficacy to the fundamental constituents of the world. They are by themselves able to generate all the variety to be found in the world. The high-level descriptions of the world we find so useful are just that: useful – indispensably useful even – epistemological aids to beings with a certain explanatory and predictive agenda who are smart enough to be able to pick out and/or formulate such descriptions[10]. It would, for example, be insane to forbid gunnery officers or pilots from thinking in terms of the coriolis force, but that cuts no metaphysical ice.

The economy argument strikes me as one reason to accept generalized epiphenomenalism. But there are other arguments as well. Consider next,

The Screening Off Argument

1999, p. 116). But this won’t work; first-order realizers are complexes of fundamental features and thus, according to my argument, have in themselves no causal efficacy. Everything they can do in the world is entirely the work of their own constituents. Realizers are not fundamental but are themselves patterns which are picked out by their relation to pre-existing high-level patterns (such as the elements of psychology, economics, geology or whatever). The economy argument shows that there is no need to suppose that realizers as such have any efficacy; if we imagine a world in which they lack efficacy the world proceeds just as well as an imagined world in which they do have efficacy (unless, of course, we enter the realm of radical emergence). Metaphysical economy then strongly suggests that we take our world to be the former world. In fact, realizers are in a sense worse off than the unrefined descriptions of high-level theory, for at least these latter have an explanatory role within their own theories whereas the realizers are epistemically inaccessible and explanatorily (as well as causally) impotent. We believe in them *because* we believe in completeness, closure and resolution but they are, in most cases, very remote from the features they are supposed to realize and can rarely take part in our detailed explanatory projects.

[10] It is indeed wonderful, and lucky for us, that the world organizes itself into reasonably well disciplined and stable high-level entities and processes. On the other hand, if it didn’t we wouldn’t be here to notice the failing.

It is often difficult to tell the difference between a cause and 'mere correlation'. Suppose we notice a correlation between, say, A and B, one in which B is a candidate for causing A. A probabilistic symptom of such a situation is the statistical relevance of B to A: $P(A|B) > P(A)$. That is, the presence of B increases the chances of obtaining A (as for example the presence of 'cigarette smoking' increases the chances of 'cancer'). One way to distinguish a mere correlate from a cause is the statistical property of 'screening off'. Essentially, the idea is to see if one can discover a feature distinct from B that accounts for the statistical relevance of B to A, and accounts for it in such a way as to undercut B's claim to efficacy. We say that C *screens off* B from A whenever

(C1) $P(A|C \& B) = P(A|C)$, but

(C2) $P(A|C \& B) \neq P(A|B)$.

We might say that, when (C1) and (C2) are met, C usurps the putative causal role of B. The test is at best a sufficient ground for denying efficacy to B, for it seems possible that $P(A|C \& B)$ might end up equal to $P(A|B)$ just by *accident*. In such a case, B might well be non-efficacious but the screening off test could not reveal this (there would be a standoff with respect to the test).

Nonetheless, the test is often effective. A classic example is an apparent link between cancer and consumption of coffee (the example is from Wesley Salmon ??). This spurious correlation resulted from not distinguishing coffee drinkers who smoke from those who do not smoke. Since coffee drinkers tend, more than non-drinkers, to be smokers a merely statistical correlation between coffee drinking and cancer arises. The screening off test reveals this since the statistics end up as follows: $P(\text{cancer}|\text{smoking} \& \text{coffee}) = P(\text{cancer}|\text{smoking}) \neq P(\text{cancer}|\text{coffee})$. The weakness of the test is nicely revealed in this example too, for it is evidently *possible* that absolutely every coffee drinker should be a smoker and vice versa.

The screening off test can be applied to the question of the causal efficacy of high-level features, with perhaps surprising results[11]. Let's begin with a toy example. Suppose I have a pair of dice. Let the low-level or micro features be the exact values of each die upon any throw (for example *getting a 2 and a 6*). A host of high-level or macro features can be defined as 'patterns' of low-level features, for example, if we take the sum of the individual outcomes, *getting an even*

---

[11] My results are contrary to an earlier attempt to use the screening off test to show that high-level features can take efficacy away from low-level features (see Brandon).

number (e.g. by rolling a 3 and a 3), *getting a prime number*, *getting a result divisible by six*, etc. Despite its simplicity the model has some suggestive attributes. There is obviously a relation of supervenience of macro upon micro features, that is, micro features *determine* macro features. But at the same time there is *multiple realizability* (there are many micro-ways to roll a prime number for example). And there is a weak analogue of non-reducibility in the sense that there is only a disjunctive characterization of macro features in terms of micro features (it is the simplicity of the example that renders the disjunctions tractable and easily comprehensible -- for contrast suppose we were rolling 10^{23} dice).

The point of the model is that micro features can *screen off* macro features from various outcomes. Consider the following probabilities. The probability of rolling a prime number is 5/12. The probability of rolling a prime *given* that one has rolled an odd number is 7/9. These are macro descriptions. Consider now one of the particular ways of rolling a prime number, say rolling a 3 and a 2. The probability of rolling a prime *given* a roll of a 3 and a 2 is, of course 1. Thus (C2) of the screening off relation is fulfilled. It is also evident that (C1) is met. Thus rolling a 3 and a 2 screens off rolling an odd number from obtaining a prime number. Since it is impossible to get an odd number except via some micro realization or other, and since each such micro realization will screen off the macro feature we should conclude that it is the micro features that carry 'efficacy'. I need the preceding scare-quotes since there is no causation within my purely formal model, but that is not a serious defect. In fact, we could add genuine causation simply by imagining a 'prime detector' which scans the dice after they are rolled and outputs either a '1' (for prime) or '0' (for non-prime). If we assume for simplicity that the detector is perfectly reliable (probability of proper detection is 1) then the probabilities of the simple example carry over directly and we get genuine causal screening off of the macro features by the micro features.

Notice that the probabilities of the outcomes contingent upon macro features are determined by some – more or less complicated – mathematical function of the probabilities of those outcomes contingent upon all the possible realizing micro features. This function, as it were, destroys the detailed information available in the micro features, changing probabilities that are all 1 or 0 to the 'smudged out' values characteristic of the macro feature probabilities (in our example, a bunch of 1s and 0s transform into 7/9).

We can apply the lessons of this simple model to more interesting cases. Consider whether basic physical features screen off chemical features. Take as an example, the chemical feature of acidity or 'being an acid'. This is multiply realizable in quite distinct physical micro features (viz. the physical realization of HCl versus that of H_2SO_4). It seems fairly clear that micro structure will screen off macro structure relative to the effects of acids. We might claim that substance S's dissolving in X was caused by X's acidity, but there are acids too weak to dissolve S so the probability of dissolution is not 1, but each particular realization state is of an acid either strong enough to dissolve S or not, and each realization guarantees we have an acid, so screening off will occur in a way quite analogous to our simple dice example.

The most interesting case is that of mentality. The same lessons can be drawn. We might claim that Fred raised his arm because he believed he knew the answer. Of course, believing that one knows the answer does not guarantee that one will raise one's arm. That is, the probability of any behaviour, given an explicatory mental state is not 1. But the micro structure in which beliefs are realized -- the specific fundamental physical features of subject and (more or less local) environment -- can plausibly be thought either to guarantee that Fred's arm goes up or to guarantee that it does not[12]. And since the realizing state by hypothesis subvenes the psychological state, we fulfill the two conditions of screening off. We expect that the probability of behaviour given a more or less gross psychological characterization is a mathematical function of the probabilities of that behaviour contingent upon all the possible realizing states. The latter probabilities will, in general, be different from the former as information is smudged out or lost in the macro psychological state relative to the determinate physical realization that will obtain in every instance of a psychological state[13].

---

[12] Even if genuine indeterminacies lurk within Fred's cognitive machinery we would expect the probabilities to approach one or zero, and to differ from the probabilities contingent upon the gross psychological characterization.

[13] Steven Yablo has pointed (see Yablo 19??) out that the debate on causation looks different from a point of view that takes the determinable/determinate distinction as its model of inter-level connection instead of the realization viewpoint. To take a common example, if a bull gets mad because it sees a red cape, the causal efficacy of the redness of the cape does not seem – at first glance – to be undercut by noting that the cape must actually have been some determinate *shade* of red. Although – despite the ingenuity of the idea – I don't think that it is very plausible to regard the physical realization states as determinates of a mentalistic determinable, the screening off argument seems to work against the causal efficacy of determinables versus their determinates. Regard the bull as a 'red-detector' (I hear that bulls are really colour blind, but let's not spoil a

Another way to approach the screening off of macro by micro features is in terms of Dennett's *stances* (see Dennett 19??). Recall that Dennett outlines three stances, that is, view-points from which to make predictions or explanations of behaviour (broadly construed to include more than that of biological organisms). We have first the 'intentional stance' in which the ascription of, most basically, beliefs and desires serves to predict and explain behaviour. The intentional stance is undoubtedly highly successful and widely applicable throughout, at least, the animal kingdom and certain subsets of the machine kingdom. But sometimes the application of the intentional stance to a system for which it is generally successful fails in particular cases. If there is no intentional stance (i.e. psychological) re-interpretation that plausibly accounts for the failure we must descend to a lower stance, in the first instance to the 'design stance'.

The 'design stance' provides the sub-psychological details of the implementation of the system's psychology, but in functional terms (Dennett, along with most cognitive scientists, typically imagines the psychology is implemented by a system of 'black boxes' with a variety of sub-psychological functions, viz. short term memory, visual memory buffers, edge detectors, etc.). Certain features of the design of a system can account for why intentional explanations fail. For example, it may be that unavoidable resource constraints force the design to be less than fully rational. For example, it will surely fail to deduce *all* the consequences of its current information, and the way it winnows out the wheat of useful information from the chaff of irrelevant noise will always be susceptible to exploitation in particular cases. The details of how relevance is assigned may lead to bizarre, psychologically inexplicable behaviour in certain circumstances (a famous example is the *sphex* wasp's seemingly purely mechanical nest provisioning behaviour, see ??).

---

traditional example with pedantry). Detectors are more or less efficient; that is the probability of detection varies with the shade presented. Thus the probability of the bull getting mad *given* that it is presented with red is not 1, and is in fact determined by the particular probabilities of getting mad given presentation of various shades. Thus perhaps the bull is very good at detecting crimson but weak at vermillion. Since both crimson and vermillion are shades of red, we will fulfill the two conditions for screening off. We ought to attribute efficacy to the determinates (the shades or red) over the determinable (red). The fact that bull, so to speak, doesn't care about or even know anything about these determinates is irrelevant to assigning efficacy. The bull's behaviour is governed by them, and correlations between colours and behaviours stem from the power of the determinates rather than the diffuse and derivative 'power' (as expressed in the probabilistic correlations) of the determinable. (Of course, we can go further, and demote the shades as well in favour of the ultimate, fundamental physical determinants of the bull's behaviour.)

But there will be failures of design stance predictions and explanations too. A random cosmic ray may chance to rewrite a bit in your computer's memory leading to behaviour that is inexplicable from the design stance (even at the level of the individual memory chip there is a design stance, and the action of the cosmic ray forces us out of it). Intrusions of brute physical malfunction are by definition inexplicable from the design stance but can never be completely 'designed away'[14]. In such cases, we must descend to what Dennett calls the 'physical stance' to account for the failure of the system to abide by its design. The possibility of the failure of the intentional stance as well as the design stance will be reflected in the probabilities of behaviour contingent upon intentional states or design states. The physical stance – at the level of fundamental physical reality – will however provide the rock bottom probabilities of behaviour contingent upon particular physical realizations of either intentional or design states. In general, these probabilities will differ from those at the intentional or design level, for their maximal information content will be smudged out across the very large disjunctive characterizations – from the physical point of view – of the intentional or design level states. Thus the existence of and relationships between the three levels results in the screening off of high-level features by the low-level features is just the way I've outlined above. The low-level, physical stance features are what generate the correlations that show up at the higher levels and it is they that deserve to be granted efficacy. This screening off reveals that no efficacy needs to be assigned to the high-level features to obtain the correlations observed at those higher levels.

Thus it seems that a good probabilistic test for efficacy further suggests that efficacy resides exclusively at the most basic level of physical reality. Yet another argument serves to reinforce this conclusion further.

The Abstraction Argument

A good high-level description manages to explain and predict the behaviour of very complex physical systems while somehow ignoring most of the complexity. Perhaps it is not a

---

[14] As mentioned above, it is our 'good fortune' to live in a world where it is relatively easy to produce systems that abide by design to an extremely good approximation (one can give an anthropic argument to the effect that if this were not so we would not be here to notice the failing). This goes both for our own technological designs but also throughout the hierarchy of the sciences, which, in accordance with completeness, closure and, especially, resolution, can all be seen as levels of *design* relative to fundamental physics, though the imputation of design is of course entirely metaphorical, from chemistry through biology to psychology.

necessary truth, not even a nomologically necessary truth, that fundamental physical features should conspire to constitute well behaved structures that obey laws which abstract away from the details of those very structures (see Kaufmann ...). But we live in a world where such behaviour is ubiquitous. That is what makes high-level theories (and theorizers) so much as possible.

The way that high-level theory abstracts away from the physical details undercuts any claim that high-level features have causal efficacy, as opposed to explanatory or predictive usefulness. Begin with a simple and famous example. While developing his theory of universal gravitation, Newton postulated that every particle of matter attracts every other particle with a gravitational force that obeys his famous inverse-square equation. But it would be obviously impossible to calculate any behaviour resulting from gravitational forces if one faced an extreme version of the many-body problem before one could even begin. And it seems that one does face such a problem for the Earth, and all other celestial bodies for which we might be interested in computing orbits, are manifestly made of a very large number of material particles (evidently held together by gravity itself among other possible forces). However, the structure of the gravitational force law and the newly invented calculus enabled Newton to *prove* that a spherical body will have a gravitational field (outside of itself) exactly *as if* all the matter of the body were concentrated at a single point – the geometric centre of the body. This is a beautiful piece of mathematics and makes calculating the orbits of the planets and the trajectories of rockets at least possible if not easy. We can now treat a planet (at least from the point of view of gross gravitational studies) as a single gravitating 'point'. We have developed a high-level description of the gravitational field of material bodies that abstracts away from the details of their material composition, to our great advantage. But would anyone want to deny that the efficacy of gravitation still resides entirely within the mass of each particle that makes up the body in question? Although Newton's theorem would provide an irresistible shortcut for any real world simulation, our imaginary superduper simulation will reveal the effects of gravitation while blithely ignoring the fact that, so to speak, all of a body's gravity can be viewed as emanating from its geometric centre (which viewpoint works even if, by chance, the body has a hollow core and there is no matter at all at the 'gravitational centre'). Such a pure mathematical abstraction cannot participate in the go of the world however useful it might be in the business of organizing our view of the world.

Another example to the same effect is that of the temperature and pressure of a gas. The reduction of thermodynamics to statistical mechanics is one of the glories of physical science and we can see it as the discovery of *how* a particular set of high-level features are grounded in lower level physical entities. We now know how temperature and pressure are realized in a gas. The temperature of a gas (at equilibrium) is the average kinetic energy of the particles which constitute the gas. The pressure of a gas is the average force per unit area which these particles exert on the gas's container. But *averages* are mathematical abstractions that in themselves cannot cause anything at all. If one is inclined to think otherwise, consider this example: a demographer might say that wages will go up in the near future since the average family size fell 20 odd years ago (and so now relatively fewer new workers are available). There is not the slightest reason to think that 'average family size' can, let alone, *does* cause things although I think we easily understand the explanation to which such statistical shorthand points[15]. By their very nature, pressure or temperature are not one whit less statistical 'fictions' than is average family size. The ascription of causal efficacy to, say, pressure is only a *façon de parler*, a useful shorthand for the genuine efficacy of the myriad of micro-events that constitute 'pressure-phenomena'. It is entirely correct to use the overworked phrase, and say that pressure is *nothing but* the concerted actions of the countless particles that make up a gas. Within our thought experimental simulation, pressure and temperature will emerge without having been coded into the simulation program. In this case, our high-level features *have* to be benignly emergent by their very nature for they are simply – by definition – mathematical abstractions of certain properties of the low-level features which realize them. As such, no genuine efficacy can be granted to them[16].

Now, we don't know what sort of realization function will be appropriate for mental

---

[15]One possible snare: the *conscious apprehension* of 'average family size' appears able to cause things but examples like these are – if they are examples of efficacy of any kind – examples of the efficacy of representational states of mind, not of the efficacy of what is represented. Thoughts about unicorns have their effects, but admitting this does not concede any causal powers to unicorns.

[16]I have to note some remarks of Richard Lewontin. Though he made them in a debate about the nature of IQ and IQ testing, I think the point ought to be generalized: 'It is important to point out that the distinction between mental constructs and natural attributes is more than a philosophical quibble, even when those constructs are based on physical measurements. Averages are not inherited; they are not subject to natural selection; they are not physical causes of any events' (NYRB, Feb. 4, 1982).

states. It is possible that the way that thermodynamical properties emerge out of the molecular dance may have more relevance to psychology than mere metaphor. There are deep analogies between thermodynamics and the dynamics of neural networks (see for example Churchland and Sejnowski 1992), and if the latter underlie psychology then some psychological properties may be surprisingly closely analogous to thermodynamical properties. Be that as it may, whatever mathematical structure underlies the transition from neural to mental states, it will be a mathematical abstraction from the underlying properties. (And the underlying neural structures will be an abstraction from the more fundamental physical structures that realize *them*.) Insofar as mental states are seen as such mathematical abstractions they cannot be granted causal efficacy. They will be mathematically convenient ways of thinking about the mass actions of the fundamental physical constituents, just as temperature is a mathematically convenient way to work with the mass action of the molecules that make up a gas.

I believe that the general structure of the SPW, plus the three arguments just advanced (which merely draw out or highlight certain consequences of the scientific picture) make a strong case for generalized epiphenomenalism. It remains to consider if the doctrine has any significant consequences for our view of the world and our place in it.

The spectre of classical epiphenomenalism is frightening since it turns us into mere spectators, utterly unable to affect the world no matter how much we might desire to. Classical epiphenomenalism makes our position in the world analogous to that of someone on a ride in Disneyland, able to watch what's happening and feel fear or elation in accord with those events, but unable to get out of the vehicle in which they are inexorably carried along and grapple with the world itself. Generalized epiphenomenalism does not seem quite so frightening. It grants no more, but also no less, efficacy to our mental states than it does to any other high-level feature of the world. We can understand perfectly well the place of high-level phenomena in the scientific picture. Such phenomena are integrated into the causal structure of the world in virtue of their being realized by structures that are indisputably efficacious (structures that really are the 'drivers' of the world or that possess the go of the world). So even if high-level features are, so to speak, metaphysically epiphenomenal, we can still coherently make a distinction between genuine and merely apparent causes at high levels of descriptions. We discover that coffee drinking does *not* cause cancer, but not because it, like every other high-level feature, is epiphenomenal[17], but because the realizers of coffee drinking do not drive the world into a state which can be given the high-level description: cancer; we find the opposite story for smoking.

The SPW is austerely beautiful, and seems to find everything it needs to construct the entire structure of the world, in all its complexity, in a handful of elementary features that constitute those structures and drive their interactions entirely ‘from below’. And it seems that just as high-level theories provide useful and in fact indispensable ‘abbreviations’ for certain well disciplined complexes of elementary aspects of the world, so too we could speak of ‘high-level causation’ as a useful abbreviation for the twin stories of *resolution* and *basic causation*.

That would be a happy ending to my story, if it was the end. Unfortunately it is not. For I believe that it is *impossible* to regard mental states as high-level features of the world in the same way that chemical, geological, etc. states are high-level features. Thus it is impossible to regard the epiphenomenalism of the mental as unproblematic in the way I’ve conceded that the epiphenomenalism of high-level features in general is unproblematic. The problem, baldly stated, can be put in the form of a dilemma. From within the confines of the SPW, one can be either an eliminativist or a perspectivalist about high-level causation. The first horn is decidedly unattractive (and I’ve been quietly suppressing it so far) since it leads directly to classical epiphenomenalism. The problem with perspectivalism is more subtle: it is that the high-level features for which generalized epiphenomenalism can be regarded as a harmless metaphysical curiosity are essentially *mind-dependent*. High-level features are elements of *patterns*, and patterns are, so to speak, invisible to the world. Their relationships, including those we like to call relations of high-level cause and effect, are only visible to conscious creatures who have invented those ways of looking at the world, and they serve only as aids to explanation, prediction and understanding for those beings rather than metaphysical go.

To say that high-level features are mind-dependent is to say that, in some sense, if there were no minds there would be no such features. This is not to say that a world devoid of mind would not be such as to be able to manifest these features to appropriately prepared minds. It is very likely that there was time when there were no minds. Though it is impossible to say when the

[17] Imagine the chances of success of tobacco company executives arguing that generalized epiphenomenalism shows that smoking not only does not but *could not* cause cancer.

first minds appeared in the universe, the *early* universe (maybe the first half-billion years or so) seems pretty clearly to preclude the kinds of organization necessary to underpin thought and experience. It would be odd to say that chemical processes were not occurring at that time. While true, this does not undercut the mind-dependency of chemical processes as high-level features. For to say that chemical processes were occurring before the advent of mind is just to say that the world was running in ways such that the application of chemical concepts would have been explanatorily fruitful (had there been minds) before minds had in fact appeared on the scene. The world had, of course, no sense of chemistry and no need of it just because chemistry is, by its high-level nature, epiphenomenal.

To put the point another way. It is impossible to understand the nature of high-level features without a prior understanding of the epistemic needs and goals of conscious thinkers. They are the sort of thing partly constituted out of those epistemic needs and goals. Consider an example. The stars visible from the Earth (say from the northern hemisphere) form literal patterns which are apparent to anyone (a little preparation or guidance helps one to see the patterns favoured by one's own culture). These constellations are useful in a variety of ways, of which the most basic, and ancient, is probably navigation. It is, I hope, abundantly clear that these patterns are mind-dependent, even though the configurations of the stars might have been the same as it has been throughout recent human history at any time in the past, even long before any consciousness of any kind was around to notice the stars (as a matter of fact, constellations are relatively ephemeral but that's irrelevant to the example). One might object that constellations aren't 'real'. But why not? In support of their unreality, the most natural insult to heap upon the constellations is that they don't *do anything* save via the mediation of mind. This will not distinguish chemistry from constellations; the point of generalized epiphenomenalism is to show that high-level features are uniformly without causal efficacy as such. Chemistry only *looks* efficacious from the point of view of mind prepared to notice and exploit 'chemical patterns'. In the loose sense countenanced above, in which efficacy is granted to a high-level feature via the efficacy of the constituents of its low-level realizers, the constellations again come out on a par with chemistry – the constituents of the low-level realizers of the constellations have lots of effects, including effects on (the constituents of the realizers of) human senses.

Now, if patterns are mind-dependent and all high-level features are patterns, then minds are mind-dependent. There is an obvious trivial (and irrelevant) sense in which this is true: if there were no minds there would be no minds. The serious problem is that the SPW requires minds in order to integrate high-level features into a world that obeys completeness, closure and resolution. But minds appear to be high-level features themselves. Therefore, there is one high-level feature (at least) that cannot be integrated into the SPW in the usual way. If we try to regard minds as a 'normal' high-level feature of the world, instead of integration with the SPW, we simply are left with 'another' mind which serves as the viewpoint from which the target mind appears as a pattern. This is either a vicious regress or an unacceptable circularity.

It turns out that the problem with generalized epiphenomenalism is not the epiphenomenalism, which can be re-interpreted for almost all high-level features in a way which preserves our common sense ascriptions of causal efficacy. The real problem is far more serious. After all, we could bit the bullet of mentalistic epiphenomenalism; it is a coherent, if bizarre and implausible, doctrine, and it *has* been accepted by serious thinkers (for example, Thomas Huxley and Frank Jackson). Generalized epiphenomenalism reveals, however, that the SPW, and in particular, the scientific picture of the metaphysics of high-level features, is *incoherent*. It wants to regard mind as a high-level feature of reality, but then finds it needs mind itself to make sense of its notion of high-level features. It does not seem possible to evade this problem, for the only place for mind within the current SPW is as a high-level feature[18]. But in order for it take such a place, there must be a perspective from which the 'patterns of mind' are apparent amidst the complex turmoil of the elementary features of the world. Such a viewpoint is itself an aspect of mind, and so we have failed to understand mentality in general as a high-level feature.

The SPW, as expressed in the doctrines of completeness, closure and resolution remains very attractive. A diagnosis of exactly what is at fault with the picture would be a difficult task far beyond the aspirations of this already overlong paper. I'll rest content with showing that such a diagnosis is urgently needed.

William Seager
University of Toronto at Scarborough

---

[18] Any other option leads to the fundamental transformation of the SPW, either via the introduction of radical emergence or perhaps through an acceptance of panpsychism. Maybe the mind could be an *elementary* feature of the world contributing some essentially mentalistic go to the world. Needless to say, this too would be a radical revision of our the SPW.